Participants

527 participants originally completed the survey, and we excluded any participants that alluded to the hypothesis in the suspicion check by responding that the study was about race and/or race and voice perceptions ($N = 20$; 3.8% of total sample). The final sample consisted of 507 (278 Women, 229 Men) participants from Amazon Mechanical Turk (see Table A1 for demographic information). Ages ranged between 19 and 82 years (_M___age_= 40.07, $SD = 13.26$). We included participants based upon the following criteria: (a) adults on (b) Amazon Mechanical Turk) (c) born and currently residing in the US (d) have had 90% or greater of their previous HITs approved, and (e) have a device with audio capabilities. We excluded Black individuals during the pre-screening process, since we are primarily interested in understanding the factors that affect threat and leadership perceptions of Black men, and group membership may differentially affect these perceptions.

Design

The study was a 2X2 within-subjects design with two independent variables: voice pitch (high or low) and race (White or Black names). Each of the four conditions was counterbalanced. Names and individual voices were randomly assigned to participants without repeat. This ensured that individuals would not listen to a high and low voice that resulted from the same original voice.

Procedure

Participants were recruited from Amazon Mechanical Turk by posting a HIT (human intelligence task) on the site. They were told that they would listen to a participant that previously provided their recording and took a "series of character trait and performance tests," which would then be compared to the participants' ratings to assess the accuracy of their perceptions. Upon being assigned to a recording, they learned the participant's name, and were provided other information about the recording (i.e., location, date) to make the design less conspicuous. Then, they listened to the participants' recording by clicking on the Soundcloud file embedded in the survey.

All of the names were randomly assigned to correspond to the high-pitched or the low-pitched conditions. The presentation of the four names for the recordings was randomized and counterbalanced across participants. We verified that the randomization worked by comparing the number of participants that were assigned to each condition, which were relatively uniform. The four conditions (Black name high pitch, Black name low pitch, White name high pitch, White name low pitch), were equally presented first, second, third, and fourth (see Table A2).

They were asked to assess the participant's character based upon their voice using a series of 100-point slider scale questions (i.e., trustworthiness, dominance, threateningness), which served as our measures of perceived trustworthiness, perceived dominance, and perceived threat, respectively. The presentation of the scale items was counterbalanced for each participant and within each condition. Additionally, we asked them to rate the individuals in the recording on various traits that were independently rated as important for leaders on 100-point slider scale items. Finally, they indicated their preferences for engaging in different types of interdependent relationships with the people in the recording on 100-point slider scales. Participants could listen to the recordings as frequently as they desired before rating the voices. They completed demographic questions and indicated what they thought the study was about as a suspicion check. After participants completed the suspicion check, we determined whether the manipulation of the names elicited perceptions of the race of the recorded individuals through a series of manipulation check questions. First, we created a name attention check score based upon whether the participants remembered the names of the people in the recordings. The participants were presented with a list of eight names, four of which were included in the study. Every time they correctly identified a name that was presented to them during the study, they received a point, for a total name recall score of four points ($M = 3.09$, $SD = .979$). If they incorrectly selected a name that was not presented to them, they did not receive a point. On the name attention check,

participants recalled the Black names with greater accuracy (68.34%) than White names (59.69%). Also, they remembered the conditions presented first (69.09%), second (70.80%), and fourth (76.30%) better than they remembered the condition presented third (39.88%). Participants were asked how many people in the recordings they thought were White or Black (see Figures A1 and A2). Finally, we asked participants to rate the likelihood that people with the names used in the study would be White or Black (see Table A3). A debriefing page explaining the true purposes of the study and the logic behind the deception was provided before payment. Participants were paid $1.00 for their participation.

Materials and measures

Voice stimuli. For the voice stimuli, we recorded the voices of eight White men between 18-30 years of age in Audacity using the Zoom H4N Handy Recorder with a sampling rate of 44.1kHz. The men quoted the first sentence of the Rainbow Passage (e.g., "When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow") (Fairbanks 1960). At the end of each sentence, the men read a randomly assigned identification number provided by the researchers. The four-digit identification numbers were created randomly, and participants were required to enter the identification number as a means of verifying that they were listening to the recordings.

After the recording sessions, each voice was manipulated to have a higher or lower pitch in Praat (Version 6.0.36) (Boersma and Heuven 2001), which served as our manipulation of threat potential through the voice. We followed the standard methods in voice research by raising and lowering each voice by 0.5 equivalent rectangular bandwidths (ERBs) using the Pitch-Synchronous Overlap Add tool in Praat, which produces a shift in perceived pitch of approximately 20 Hz in either direction (e.g., Apicella and Feinberg 2009; Klofstad, Anderson, and Peters 2012; Tigue et al. 2012; Vukovic et al. 2010). We set the pitch floor to 70 Hz and the pitch ceiling to 250 Hz, which has been validated as an appropriate range for male voices (Vogel et al. 2009). Many researchers manipulate ERB instead of Hertz because a change in pitch is perceived differently depending upon the original pitch that was manipulated, since there is a logarithmic relationship between actual pitch and perceived pitch (Stevens, 1998). Also, the ERB manipulations will not affect other acoustic characteristics of the recording (e.g., rate, intensity) (Feinberg et al. 2005). Since each of the voices was raised and lowered in pitch, there were a total of sixteen manipulated recordings. We checked the manipulation by comparing the mean pitch for the original voices (M = 104.37, SD = 14.09) to the lower manipulations (M = 90.21, SD = 9.79) and the higher manipulations (M = 121.30, SD = 17.28). All of the manipulated files were uploaded to separate Soundcloud links and embedded in the survey.

Names for race manipulation. To manipulate perceptions of race, we used four names that are typically associated with Black people (i.e., Tyrone, Keyshawn, Deshawn, Terrell) and four names that are typically associated with White people (i.e., Scott, Brad, Brett, and Logan) (Gaddis 2017). Names were presented before the participants listened to the voice recording. Each name was chosen based upon the criteria that 90% or more of raters from Gaddis (2017) thought that the individual was either Black or White when they were asked about their perceptions of the person's race based upon their name.

Perceived leadership ability. We recruited 55 participants on Amazon Mechanical Turk to serve as independent raters for identifying the leadership traits used in the experiment. We provided them with a list of fifteen traits from which they could select what they considered most valuable for successful leaders of businesses and companies (e.g., drive, creativity, confidence) (Kirkpatrick and Locke 1991). We selected the traits that were ranked, on average, in the 30th percentile of responses (where 1 is considered the most important trait for a leader). The traits that were selected for the leadership composite score based upon these criteria were intelligence, effective communication, confidence, and problem-solving ability, which were rated by participants in the final study using 100-point slider scale items.

To create the leadership ability composite, we averaged participants' ratings of the individual in the recording on the four traits. Higher scores denote greater perceived leadership ability. The measure had high internal consistency across participants in the final sample (alpha = .91; averaged across all conditions).

Perceived threat, trustworthiness, and dominance. Single questions were used to elicit perceived

threat, trustworthiness, and dominance. Participants responded using a 100-point slider scale.

Apicella, Coren L., and D. R Feinberg. 2009. "Voice pitch alters mate-choice-relevant perception in hunter-gatherers." *Proceedings of the Royal Society B: Biological Sciences* 276: 1077–82. https://doi.org/10.1098/rspb.2008.1542.

Boersma, Paul, and Vincent van Heuven. 2001. "Speak and unSpeak with Praat." *Glot International* 5 (9-10): 341–47.

Fairbanks, Grant. 1960. *Voice and articulation drillbook*. New York, NY: Harper.

Feinberg, D. R., B. C. Jones, A. C. Little, D. M. Burt, and D. I. Perrett. 2005. "Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices." *Animal Behaviour* 69 (3): 561–68. https://doi.org/10.1016/j.anbehav.2004.06.012.

Gaddis, S. 2017. "How Black Are Lakisha and Jamal? Racial Perceptions from Names Used in Correspondence Audit Studies." *Sociological Science* 4: 469–89. https://doi.org/10.15195/v4.a19.

Kirkpatrick, Shelley A., and Edwin A. Locke. 1991. "Leadership: do traits matter?" *Executive* 5 (2): 48–60. https://doi.org/10.5465/AME.1991.4274679.

Klofstad, C. A., R. C. Anderson, and S. Peters. 2012. "Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women." *Proceedings of the Royal Society B: Biological Sciences* 279 (1738): 2698–2704. https://doi.org/10.1098/rspb.2012.0311.

Tigue, Cara C., Diana J. Borak, Jillian J. M. O'Connor, Charles Schandl, and David R. Feinberg. 2012. "Voice pitch influences voting behavior." *Evolution and Human Behavior* 33 (3): 210–16. https://doi.org/10.1016/j.evolhumbehav.2011.09.004.

Vogel, Adam P, Paul Maruff, Peter J Snyder, and James C Mundt. 2009. "Standardization of pitch range settings in voice acoustic analysis." *Behavior Research Methods* 41 (2): 318–24. https://doi.org/10.3758/BRM.41.2.318.Standardization.

Vukovic, Jovana, Benedict C. Jones, Lisa Debruine, David R. Feinberg, Finlay G. Smith, Anthony C. Little, Lisa L. M. Welling, and Julie Main. 2010. "Women's own voice pitch predicts their preferences for masculinity in men's voices." *Behavioral Ecology* 21 (4): 767–72. https://doi.org/10.1093/beheco/arq051.